

存活分析

Survival Analysis

第二版

林建甫

Jeff Chien-Fu Lin, MD. PhD.

存活分析 / 林建甫著. - 二版. - 台北市:

雙葉書廊, 2020. 02

面: 公分

ISBN 978-957-9096-66-9 (平裝)

1. 生物統計 2. 醫學

306.13

109000884

©2020 林建甫

書名: 存活分析

著者: 林建甫

著作權與版權所有人: 林建甫

出版社: 雙葉書廊有限公司

發行人: 張福隆

地址: 台北市羅斯福路三段 269 巷 12 號 1 樓

<http://www.yehyeh.com.tw>

電話: 886-2-2368-4198

傳真: 886-2-2365-7990

出版日期: 西元 2020 年 2 月 二版一刷

再刷日期: 西元 2020 年 7 月

ISBN: 978-957-9096-66-9 (平裝)

本書著作權與版權屬於著作者所有, 受中華民國著作權法保護, 未經著作者書面同意, 不得以任何形式轉載, 複印, 翻印或重製。

作者 林建甫

Jeff Chien-Fu Lin, MD, PhD.

cflin@mail.ntpu.edu.tw

<https://jefflinmd.com>

經歷 巨匯數據科技股份有限公司 統計顧問

美國范德堡大學醫學中心 計量科學中心 生物資訊與癌症研究 訪問學者

智策市場研究顧問公司 統計顧問

台北大學 統計系 助理教授

台北榮民總醫院 生物統計顧問

台北市 中山醫院 骨科醫師

台北醫學大學 萬芳醫院 骨科醫師

高雄長庚醫院 骨科醫師

台北榮民總醫院 骨科醫師

學歷 美國密西根大學 生物統計研究所 碩士, 博士

高雄醫學大學 醫學系 醫學士

興趣 醫學統計諮詢, 臨床試驗, 基礎醫學實驗設計

存活分析, 縱向資料分析, 類別資料分析

生物資訊, 大數據分析, 醫療經濟成本效益評估

二版序

第二版的修訂與第一版相距 10 年以上,在這段時間,存活分析的論理論與應用的發展非常迅速,統計軟體發展也是有巨大的改變,然而本書第二版的寫作對象如同第一版所敘述,主要為從事資料分析的統計人員與從事臨床醫學研究的醫師,因此修訂的方向盡可能保留原有的想法與架構,以介紹紹醫學期刊常出現的存活分析應用方法為主,第二版也盡可能地修正第一版部分錯誤的文字與符號,將第一版部分章節分割且同時保留原有架構,刪減部分方程式,縮小圖片,降低行距等,避免總頁數增加過多。第二版的修訂同時增加了部分章節,包含計數過程導論,存活預測的評估方法,競爭風險與多重狀態模型,相加性危險函數模型,區間設限的迴歸模型,樣本數計算等。第二版書共 22 章,全書分成兩大部分,第 1 章至第 10 章為基礎存活分析,第 11 章至第 22 章為進階存活分析。在第一部分基礎存活分析中,包含存活分析概論,參數模型,無母數方法估計存活函數,無母數方法比較兩個或多個樣本之存活函數,Cox 比例危險模型與存活率預測,分層 Cox 模型,時間相依 Cox 模型,Cox 迴歸模型診斷等。在第二部分進階存活分析中,包含評估存活預測正確性,截略與區間設限資料分析,多變量存活分析包含群聚存活資料與復發事件資料,競爭風險與多重狀態模型,相加性危險函數模型,以及計數過程導論等。

第二版增訂的內容是在 2008–2020 年間陸續增添,內容部分來自這 10 年在工作坊或研習營中,講授事件歷史研究法與進階存活分析的內容,一些來自這些年在存活分析課程與臨床試驗課程的上課內容,以及作者參與的臨床論文中所使用的方法。本書受限於頁數無法包含所有迅速發展的存活應用,很難完全滿足所有讀者需求。

在 2019 年初,雙葉書局通知作者關於存活分析第一版書已經售完,同時詢問作者是否有改版或是再刷計畫。回想 2008 年當初作者認為第一版書到退休時都不會賣完,當時並沒有打算繼續印書或改版。在 2013 年起就有些讀者詢問是否會有改版,這是銷售量非常少的書,而且是自費出版,考量寫作時間且須親自以 $cw\text{\TeX}$ 排版,繼續印書或改版是完全不符合任何經濟效益。但是隨後仍有讀者陸續詢問,所以在第一版書

售完後, 於 2019 年初採用 POD 的方式印刷, 雖然印刷成本增加, 仍不失為一種妥協的方法。然而若要繼續印書, 身為作者仍須考慮增加一些內容, 以符合現代存活分析應用上的完整性。第二版仍然採用 POD 方式印刷, 方便作者較容易修正錯誤以及添加新的內容。最後, 無論如何修正, 受到本人有限的學術能力與時間的限制, 第二版仍多有所錯誤之處, 尚請先進與賢達不吝指教。本書的排版是利用吳聰敏教授, 吳聰慧先生, 與翁鴻翎先生開發出來的 cwTeX 系統排版軟體所完成, 謹此致謝。本書能夠付梓, 感謝雙葉書局編輯部門的協助。

這些年來在統計工作上感謝國立台北大學統計系的前輩與同事協助。在臨床工作上受到台北市萬芳醫院骨科何為斌主任, 呂才學主任, 郭宜潔主任, 以及骨科所有同仁的協助, 特此致謝, 也特別感謝台北中山醫院吳濬哲院長給予作者長期的支持與鼓勵。第二版書能夠付梓, 感謝雙葉書局編輯部門的協助。最後, 感謝在這段期間, 凤凰, 湘宜, 靖庭的犧牲與支持, 讓我有機會繼續從事跨領域的工作。

林建甫 謹誌于
台北, 2020 年 2 月

一版序

在醫學研究中, 存活時間資料為一種常見的研究資料, 例如病患從診斷出癌症到死亡之間的存活時間. 本書主要目的是討論一些常用來分析存活時間的方法, 在醫學統計中, 這些方法常稱為存活分析. 一般而言, 存活分析方法是用來研究或分析樣本所觀測到的某一段時間長度之分配. 一段時間長度通常是從一特定事件起始之起始時間點算起, 計算到某一特定事件發生的時間點為止, 通常稱為事件時間或存活時間. 事件時間的資料出現在不同領域中, 譬如醫學中的癌症存活率; 公共衛生中的死亡率; 流行病學中的生命量表; 保險統計學及人口統計學中的生命統計資料; 工程學中的可靠度分析; 社會學中的事件歷史分析, 研究結婚時間到離婚時間, 到再婚時間, 人口居住時間與流動時間; 商業研究中, 市場研究之消費者對特定商品購買時間, 客戶忠誠度的時間; 公司企業的存活時間; 經濟研究中的失業, 從就業時間到失業時間, 到再就業時間等; 法學研究中, 犯罪嫌疑人從犯罪時間到被捕時間, 犯罪嫌疑人從被捕時間到起訴時間, 從起訴時間到定罪時間, 從假釋時間到再犯時間等.

在醫學研究中, 常常收集病患從診斷出癌症到死亡之間的存活時間, 但是, 並非所有進入研究的觀察對象, 其被追蹤的時間都足夠長, 所以研究者並不能夠觀察到所有研究對象最後的結果, 及所有發生死亡事件時間點. 這樣的資料通常稱為設限資料. 此外, 病患常常具有一些危險因子, 進而影響存活的機會; 但是, 許多影響事件之發生或影響存活時間長短的危險因子, 是在追蹤的時間內才發生, 稱為時間相依危險因子, 這樣的存活時間資料, 必須要有特殊的統計方法來分析. 這些特殊的統計方法, 通稱為存活分析, 這是一個籠統定義的統計名詞. 存活分析相關書籍中, 有許多可做為參考, 如 Lawless (2002) 對可靠度分析之參數模型有詳盡的討論. Cox 與 Oakes (1984) 則是以精簡的手筆, 觸及各項存活分析議題. Collett (2003) 討論單變量存活分析, 以應用為主. Hosmer 與 Lemeshow (1999) 討論單變量存活分析, 有一章討論多變量存活分析, 以應用為主. Klein 與 Moeschberger (2003) 主要討論單變量存活分析, 有一章討論多變量存活分析, 雖以應用為主, 但對存活理論有深入的介紹. Kalbfleisch 與

Prentice (2002) 對單變量與多變量存活分析都有討論，應用與理論並重，但較為艱難。Fleming 與 Harrington (1991)，以及 Andersen 等 (1993) 則是以計數過程為架構，說明存活分析之統計理論。Therneau 與 Grambsch (2000) 討論 Cox 延伸模型與多變量存活分析。Hougaard (2000) 則專門討論多變量存活分析。但這些書籍對長期從事臨床醫學研究的醫師，較難以一一閱讀與理解。

本書主要寫作對象為從事資料分析的統計人員與從事臨床醫學研究的醫師，本書所使用的數學內容僅包含加法、乘法，與簡單之微積分，這些內容等同於在台灣的高中與大一的基礎數學程度。閱讀本書的統計專業人員、統計系研究生或大學生，最好具備數理統計、迴歸分析、變異數分析等知識，部分章節則需一些類別資料分析與縱向資料分析等基本知識。至於從事臨床醫學研究的醫師閱讀本書時，閱讀本書時，以閱讀文字為主，閱讀方程式演算過程為輔，主要須了解書中文字與主要演算方程式之結果，與方程式背後在資料分析實務上之意義，對每一種分析方法的使用時機與其優缺點，這樣則會有助於判讀醫學文獻與選擇適當的分析方式分析手中的研究資料。

本書共 14 章，全書分成兩大部分，第 1 章至第 9 章為單變量存活時間分析，第 10 章至第 14 章為多變量存活時間分析。為了每一章節之完整性，因此少數之部份內容會有重覆之討論。在第一部分單變量存活時間分析中，第 1 章為存活分析概論，簡介書中常使用的存活資料。第 2 章說明存活函數危險函數，設限與截略定義，存活分析常見的基本假設，以及建構概似函數。第 3 章以參數模型進行存活分析，包含加速失敗時間模型。第 4 章以無母數方法估計存活函數，討論生命量表，Kaplan-Meier 估計式，Nelson-Aalen 估計式。第 5 章以無母數方法比較兩個或多個樣本之存活函數，包含 log-rank 檢定，Wilcoxon 檢定，最後討論分層檢定。第 6 章討論 Cox 比例危險模型，這是存活分析中最重要的一部份，解釋時間固定共變數之參數估計，基線危險函數與存活函數估計，模型選擇，並說明 Cox 比例危險模型在醫學上的應用。第 7 章討論常見的 Cox 延伸模型，如分層 Cox 模型，時間相依共變數與時間相依 Cox 模型，及時間相依 Cox 模型在移植醫學上的應用。第 8 章則討論 Cox 遷歸模型之診斷，如何利

用各種殘差, 檢定比例危險之假設, 或尋找最佳函數型式. 第 9 章介紹截略資料與區間設限資料的存活分析. 在第二部分多變量存活時間分析中, 第 10 章為多變量存活分析簡介, 說明多變量存活時間與單變量存活時間差異. 第 11 章討論群聚存活資料之邊際模型與分層模型. 第 12 章則討論群聚存活資料之脆弱模型與隨機效應的觀念. 第 13 章討論復發事件資料常見之 3 種廣義邊際模型. 第 14 章則討論復發事件資料之脆弱模型.

本書以應用在醫學資料分析為主, 對於單變量或多變量存活分析方法中, 若現今常見之統計軟體可以執行之分析方法, 則儘可能提出討論. 本書中所有的例題分析與圖表建構, 皆是以使用 *R* 或是 **SAS** 統計軟體執行運算, 使用這 2 種統計軟體主要是因為作者的偏好與熟悉度, 本書中多數的例題分析與圖表建構, 也可以使用其它統計軟體, 如 STATA, S-Plus, SPSS 等執行運算. 本書如同其它書籍, 不可能討論所有存活分析的議題, 在本書中所割捨的議題, 包含相加性危險模型, 多重階段模型, 競爭風險模型, 以群體癌症登記為基礎之相對存活分析, 貝氏存活分析, 以及計數過程之理論等. 希望能在下次改版中加入或另外撰寫專書討論. 最後, 因本人才疏學淺, 以致本書有所錯誤或疏漏, 尚請諸位先進與賢達不吝指教.

本書是作者在 2001-2007 年之間寫作, 以電子書型式, 供學生使用. 本書的起源約在 2001 年秋季, 作者在台北大學統計系擔任兼任教師, 講授存活分析課程, 當時台灣並未大量進口適合作為碩士研究生之教科書, 因此自行撰寫教材, 講義原稿以書本型式寫作, 內容包括本書第 1 章至第 9 章中的存活理論部分, 隔年並將大部份內容改寫與修訂. 在 2003-2005 年, 因改授其它課程而中斷寫作. 在 2006 年再次講授存活分析課程, 因此也再度改寫大部分內容, 在 2006-2007 年, 對第 1 章至第 9 章大多數之分析方法, 加入實例解說. 另外再加寫本書第 10 章至第 14 章理論與實例, 最後在 2008 年初重新排版, 整理成書.

作者長期從事醫學統計諮詢工作與期刊論文的統計審核, 存活分析常見之主要錯誤分別為: 研究者不知其研究資料為存活事件時間資料, 研究者不知其研究資料中的共變數為時間相依共變數, 研究者不知其研究資料中的存活事件時間為多變量事件

時間。作者嘗試以個人臨床醫學經驗結合醫學統計諮詢經驗，以雙方都能了解的語言與文字，將存活分析之統計方法整理成書。希望藉由本書之寫作，在統計人員與臨床醫師之間，提供一個溝通的橋樑，並共同分享分析存活資料的經驗與樂趣。

人生充滿著意外與驚喜，在偶然的機緣下，進入到醫學與統計的世界，先後受到醫學與統計兩個學門的師長，悉心指導，感激在心。本書的完成，要感謝美國密西根大學生物統計系給我機會，進入到統計的世界。首先感謝引導我進入存活分析的領域，也是我的指導教授 Robert A. Wolfe 博士；感謝在求學期間，提供我助學金的教授 Richard G. Cornell 博士，Judith Bromberg 博士，M. Anthony Shorck 博士；以及讓我有機會進入生物統計系就學的教授 Michael Boehnke 博士。在 1995-2004 年專職從事臨床骨科工作時期，感謝台北榮民總醫院骨科部與高雄長庚醫院骨科許多老師們的鼓勵，以及許多同事的協助，繼續參與醫學統計的教學與諮詢。感謝國立台北大學統計系的前輩與同事，讓我有機會參與統計系所的教學與研究工作，並參與建立台北大學統計系所的醫學統計學程。本書經過多次改寫與修訂，感謝歷年來協助修訂的研究生，包括亭嫻，月麗，志學，惠瑜，湘婷，億潔，譚景，如茵，百筠等。本書的排版是利用吳聰敏教授，吳聰慧先生，與翁鴻翎先生開發出來的 cwTeX 系統排版軟體所完成，謹此致謝。本書能夠付梓，感謝雙葉書局編輯部門的協助。特別感謝台北榮民總醫院教學研究部主任何橒通教授，高雄長庚醫院院長陳肇隆教授，台北榮民總醫院骨科部主任陳天雄教授，與台北中山醫院骨科吳濬哲主任，給予作者長期的支持與鼓勵。最後，感謝在撰寫本書這段期間，鳳凰，湘宜，靖庭的犧牲與支持，讓我有機會完成本書。

林建甫 謹誌于
台北大學統計學系
2008 年 3 月

目錄

第 1 章 存活分析概論	1
1.1 粗死亡率及死亡累積發生率	2
1.2 發生速率與發生密度	5
1.3 存活資料與設限	8
1.4 存活資料的真實案例	12
1.5 癌症研究反應變數與存活分析	19
1.6 存活分析的特徵	20
1.7 存活迴歸模型/邏輯斯迴歸/布瓦松迴歸模型 的區別	23
1.8 資料分析練習	24
第 2 章 存活函數, 設限與截略	31
2.1 存活函數及危險函數	31
2.1.1 危險函數	34
2.2 離散型存活函數	36
2.3 平均存活時間與平均餘命	38
2.4 設限和截略簡介	39
2.5 右設限觀察資料	41
2.5.1 型 I 設限	42
2.5.2 型 II 設限	43

2.6	隨機設限觀察資料	44
2.7	左設限, 雙重設限, 區間設限	45
2.8	截略觀察資料	47
2.9	設限與截略資料的概似函數	49
2.10	存活分析常用之重要統計假設	51
2.11	隨機設限概似函數	52
第 3 章 參數模型存活分析		55
3.1	存活分析常見的參數機率分配	55
3.1.1	指數分配	56
3.1.2	韋伯分配	58
3.1.3	廣義 Gamma 分配	60
3.1.4	對數常態分配	61
3.1.5	對數邏輯斯分配	61
3.1.6	標準極值(最小值)分配	62
3.1.7	極值分配 – Gumbel 分配	63
3.2	參數機率模型概似函數	63
3.3	最大概似估計參數	65
3.4	Newton-Raphson 演算法	66
3.5	參數推論: 假說檢定	68
3.6	參數推論: 信賴區間	70
3.7	Delta 法估計函數的近似變異數	73
3.8	配適簡單參數模型: 案例分析	74
3.8.1	配適指數分配	75
3.8.2	配適韋伯分配	78
3.8.3	參數族之選擇	81
3.9	存活分析之迴歸模型	82
3.9.1	比例危險模型	83
3.9.2	可加性危險模型	84

3.9.3	加速失敗時間模型	84
3.9.4	韋伯分配與比例危險模型	85
3.10	加速失敗時間模型分析	86
3.11	資料分析練習	90
 第 4 章 無母數方法估計存活函數		93
4.1	生命量表	94
4.2	生命量表的種類	104
4.3	Kaplan-Meier (Product-Limit) 估計式	105
4.4	Kaplan-Meier 估計式之 Greenwood 變異數估計	110
4.5	Peterson 累積危險函估計式	114
4.6	Nelson-Aalen 估計式	115
4.7	存活函數的逐點信賴區間	117
4.8	存活函數信賴區間帶	120
4.9	存活時間中位數及百分位數估計式	122
4.10	平均存活時間估計	123
4.11	資料分析練習	124
 第 5 章 無母數方法比較存活函數		127
5.1	檢定兩組存活函數	127
5.1.1	兩組樣本存活比較與假說檢定	128
5.1.2	無母數方法比較兩組樣本之存活分配	130
5.2	權重與檢定	133
5.2.1	Log-Rank 檢定	133
5.2.2	(Gehan) Wilcoxon 檢定	134
5.2.3	Tarone-Ware 檢定	135
5.2.4	Peto-Wilcoxon 檢定	135
5.2.5	檢定之比較與選擇	137

5.3	比較多組存活函數	138
5.4	分層 Log Rank 檢定	143
5.5	樣本數計算	147
5.6	資料分析練習	161
第 6 章 Cox 比例危險模型		163
6.1	簡單 Cox PH 模型	164
6.2	Cox 比例危險複迴歸模型	167
6.2.1	共變數影響之效力調整與交互作用	169
6.2.2	多項式 Cox 模型	170
6.3	Cox 模型概似函數	171
6.3.1	完全概似函數	171
6.3.2	Cox 的概似函數: 在沒有相同事件時間的偏概似函數	172
6.4	參數估計	174
6.5	參數推論: 假說檢定	177
6.6	參數推論: 信賴區間	179
6.7	Cox PH 模型實例說明	182
6.8	Cox 概似函數: 多個相同事件時間	187
6.8.1	Breslow 估計法	189
6.8.2	Efron 估計法	191
6.8.3	Cox 估計法	192
6.8.4	Kalbfleisch 與 Prentice 估計法	193
6.9	建構 Cox PH 模型	195
6.9.1	調整潛在的干擾變數	196
6.9.2	Akaike 資訊準則和 Schwartz 的貝氏資訊準則	196
6.9.3	逐步選取方法	198
6.10	Cox 概似函數的討論	201
6.10.1	Cox 的偏概函數如同 Profile 概似函數	202
6.10.2	Cox 對排序時間的偏概似函數與邊際概似函數	203

6.10.3	從全概似函數看偏概似函數	205
6.11	資料分析練習	207
 第 7 章 Cox 模型估計存活函數		215
7.1	基線存活函數的概似函數	216
7.2	基線危險函數之估計	219
7.2.1	估計基線存活函數: 無相同事件時間	220
7.2.2	估計基線存活函數: 相同事件時間	221
7.3	統計軟體中的存活函數估計	225
7.4	資料分析練習	231
 第 8 章 分層 Cox 模型		233
8.1	分層 Cox 比例危險模型	234
8.2	檢定分層分析中平行線假設	235
8.3	資料分析練習	240
 第 9 章 時間相依共變數與 Cox 模型		241
9.1	時間相依共變數	242
9.1.1	內在與外在時間相依共變數	242
9.1.2	研究追蹤中治療更動或改變: 時間相依共變數	243
9.1.3	治療改變後才測量之時間相依共變數	244
9.1.4	在比例危險模型中定義時間相依共變數	245
9.2	時間相依共變數的 Cox 非比例危險模型	249
9.3	Cox 時間相依模型的偏概似函數	250
9.4	Cox 時間相依模型的參數估計	252
9.5	計數過程與時間相依共變數	253
9.6	史丹佛心臟移研究資料	256
9.7	Cox 時間相依模型與估計存活函數估計	264

9.8 資料分析練習	266
第 10 章 Cox 迴歸模型之診斷	
10.1 殘差	272
10.1.1 Cox-Snell 殘差	272
10.1.2 平賭殘差	273
10.1.3 Deviance 殘差	278
10.1.4 分數殘差	278
10.1.5 Schoenfeld (分數) 殘差	279
10.2 評估 Cox 模型的適合度與 Cox-Snell 殘差	280
10.3 評估一個共變數的最佳函數形式與平賭殘差	281
10.4 檢查成比例模型假設	283
10.5 檢查 Cox 模型離群值與 Deviance 殘差	286
10.6 個體個別共變數之影響值	287
10.7 Cox 模型診斷實例	288
10.8 資料分析練習	298
第 11 章 存活預測評估	
11.1 固定存活時間預測與交叉驗證預測誤差	300
11.2 Harrell's C-指數	305
11.3 時間相依 ROC 分析	306
11.4 資料分析練習	314
第 12 章 截略資料存活分析	
12.1 左截略	321
12.2 左截略與延遲進入資料	326
12.3 左截略與時間相依的分層	328
12.4 右截略資料與存活函數	329

12.5	左設限資料與存活分析	334
第 13 章 區間設限資料分析		337
13.1	區間設限資料	338
13.2	區間設限資料與存活函數	339
13.3	區間設限資料: 存活迴歸模型	345
13.4	資料分析練習	347
第 14 章 多變量存活分析簡介		349
14.1	多重事件與多變量存活時間實例	350
14.2	群聚存活時間	353
14.3	多重復發事件時間	354
14.4	穩健的變異數估計	355
14.5	Cox 比例危險模型與多變量存活分析	356
14.6	資料建構與模型選擇	358
14.7	無序多重存活時間	359
14.8	有序多重存活時間	360
14.9	資料分析練習	362
第 15 章 群聚存活資料與邊際模型		365
15.1	邊際模型簡介	365
15.2	邊際模型之統計操作	367
15.3	邊際模型的限制	369
15.4	資料分析練習	370
第 16 章 群聚存活資料與脆弱模型		373
16.1	隨機效應與脆弱	374

16.2	群聚資料: 分層 Cox 模型	375
16.3	隨機效應與脆弱模型	378
16.3.1	脆弱模型之條件危險函數	378
16.3.2	群聚事件時間之聯合機率模型	379
16.3.3	脆弱模型之邊際與條件危險函數	380
16.4	相關性之測量	384
16.5	群聚資料隨機效應脆弱模型之檢定	385
16.6	常見之隨機效應脆弱模型	386
16.6.1	Gamma 脆弱模型	387
16.6.2	Positive Stable 脆弱模型	388
16.6.3	對數常態脆弱模型	389
16.7	脆弱模型之估計與推論	390
16.7.1	估計 Gamma 脆弱模型之 EM 演算法	390
16.7.2	PPL 演算法	393
16.7.3	Positive Stable 脆弱模型之估計與推論	395
16.8	脆弱模型實例: DRS 資料	395
16.9	脆弱模型之延伸	397
16.9.1	邊際比例危險之脆弱模型	397
16.9.2	指定錯誤的脆弱模型	398
16.9.3	檢測脆弱模型之假設	399
16.9.4	基因與脆弱模型	399
16.10	群聚脆弱模型結論	400
16.11	資料分析練習	401
	第 17 章 復發事件資料邊際模型	403
17.1	復發事件與布瓦松過程	404
17.1.1	時間軸	405
17.1.2	布瓦松過程與計數過程	406
17.1.3	布瓦松迴歸模型	406
17.1.4	邊際事件失敗率模型	407

17.2	復發事件資料: 常見邊際模型	408
17.2.1	Andersen 與 Gill 邊際模型	409
17.2.2	Wei, Lin, 與 Wiessfeld (WLW) 邊際模型	412
17.2.3	Prentice, Williams 與 Petersen (PWP) 條件模型 ..	414
17.3	模型配適與資料型式	415
17.4	膀胱癌臨床試驗: 廣義邊際模型	418
17.5	復發事件資料邊際模型摘要	423
17.6	資料分析練習	424
 第 18 章 復發事件資料脆弱模型		429
18.1	隨機效應與脆弱模型	430
18.2	復發事件: 過度分散布瓦松迴歸	431
18.3	復發事件脆弱模型與概似函數	432
18.4	復發事件脆弱模型: 膀胱癌臨床試驗	433
18.5	脆弱模型的選擇	441
18.6	資料分析練習	442
 第 19 章 競爭風險		445
19.1	特定原因危險函數	447
19.2	累積發生函數估計	450
19.3	非參數假說檢定	456
19.4	子分配危險函數	457
19.5	非參數假說檢定: 葛雷檢定	461
19.6	競爭風險: Cox 迴歸模型	463
19.7	競爭風險: Fine-Gray 模型	464
19.8	資料分析練習	468

第 20 章 多重狀態模型	471
20.1 轉移強度與轉移機率	473
20.2 非參數估計轉移強度與轉移機率	477
20.3 多重狀態迴歸模型	481
20.4 資料分析練習	489
第 21 章 可加性危險模型	493
21.1 可加性危險模型	494
21.2 可加性危險模型參數估計	495
21.3 參數假說檢定	498
21.4 可加性危險模型特例	499
21.5 資料分析練習	506
第 22 章 計數過程導論	507
22.1 機率與測度	507
22.2 隨機過程	509
22.3 計數過程與存活資料	511
22.4 Nelson-Aalen 累積危險函數估計式	513
22.5 計數過程與存活函數	516
22.6 計數過程與 Log Rank Test	516
22.7 計數過程與強度函數	517
22.8 共變數與事件時間危險函數模型	520
22.9 平賭過程	524
22.10 平賭過程與存活計數過程	529
22.11 多變量時間與計數過程	534

22.12	平賭之二次變異過程與選擇變異過程	536
22.13	連續型失敗時間: 計數過程與變異過程	545
22.14	離散型失敗時間: 計數過程與變異過程	546
22.15	向量平賭	548
22.16	平賭過程: 中央極限定理	549
22.17	危險函數與存活函數估計式漸進理論	552
22.18	Log Rank Test 漸進理論	556
22.19	Cox 模型漸近理論	557
22.20	偏概似函數過濾集	564
	參考文獻	567
	索引	577

表目錄

1.1	一個社區中肺癌的盛行率及累計發生率	4
1.2	20 的兒童血癌存活時間完整追蹤研究 (單位: 月)	9
1.3	設限資料: 觀察 50 位個體之存活時間 (單位: 月)	10
1.4	Gehan 急性白血病臨床試驗: 配對比較 6-MP 組與對照組的緩解時間	13
1.5	Gehan 急性白血病臨床試驗: 2 組比較 6-MP 組與對照組的緩解時間	14
1.6	白血病患者的白血球數與死亡時間	15
1.7	美國退伍軍人醫院肺癌臨床試驗	16
1.8	VGH 第一期非小細胞肺癌手術後存活研究: 7 位個體部分研究結果 .	17
1.9	VGH 第一期非小細胞肺癌手術後存活研究: 變數說明	17
1.10	心臟移植存活時間: 5 位病患部分資料	18
1.11	心臟移植存活時間	18
1.12	資料收集中重要的時間及日期	22
1.13	存活分析中常見的統計差異	22
1.14	德國乳癌研究: 變數說明	25
1.15	早期濾泡型淋巴癌長期追蹤資料: 變數說明	27
1.16	胃賁門癌存活分析: 變數說明	29
2.1	1990 年代人類存活率	32

2.2	設限與截略之概似函數	50
3.1	DMBA 致癌物之老鼠存活試驗	74
3.2	DMBA 致癌物之老鼠存活試驗: 配適指數分配結果	77
3.3	DMBA 致癌物之老鼠存活試驗: 配適韋伯分配結果	80
3.4	美國退伍軍人醫院肺癌臨床試驗: 韋伯分配加速失敗時間模型	88
3.5	胸膜惡性間皮瘤二期臨床試驗: 變數說明	91
3.6	胸膜惡性間皮瘤二期臨床試驗: 部分個體資料	91
4.1	存活分析: 生命量表	96
4.2	VGH 第一期非小細胞肺癌手術後存活研究的生命量表 (I)	102
4.3	VGH 第一期非小細胞肺癌手術後存活研究的生命量表 (II)	103
4.4	VGH 第一期非小細胞肺癌手術後存活研究: 在生命量表中的區間中點之危險率	104
4.5	存活分析: Kaplan-Meier 法	110
4.6	急性白血病臨床試驗中 21 位孩童接受 6-MP 治療: Kaplan-Meier 存活率函數與標準誤差估計直	113
4.7	復發惡性腦星狀細胞瘤研究: 20 位成人的死亡或設限時間 (單位: 週)	119
4.8	復發惡性腦星狀細胞瘤: Kaplan-Meier 存活率和累積危險率	119
4.9	威爾母士氏瘤資料: 變數說明	124
4.10	多發性骨髓瘤存活分析: 變數說明	126
5.1	一系列 2×2 列聯表分析: 階層 1	131
5.2	一系列 2×2 列聯表分析: 階層 i	131
5.3	一系列 2×2 列聯表分析: 階層 k	131
5.4	Mantel-Haenszel 統計量應用在一序列之 2×2 列聯表: 在時間點 t_i (階層 i) 兩組樣本存活之結果	131

5.5	Gehan 急性白血病臨床試驗: 比較 6-MP 與安慰劑治療後存活時間 .	136
5.6	一序列 $G \times 2$ 的列聯表: 群組 G 在時間點 t_i 時情況	139
5.7	F98-神經膠質瘤動物實驗: 三組治療之存活時間 (+: 設限)	142
5.8	美國退伍軍人醫院肺癌臨床試驗: 分層 log-rank 檢定	145
5.9	慢性肉芽腫病第 1 次嚴重感染時間分析 I: 變數說明	162
6.1	白血病患者白血球數與死亡時間: 單一共變數的 Cox 模型	184
6.2	白血病患者白血球數與死亡時間: 2 個共變數的 Cox 模型	186
6.3	白血病患者白血球數與死亡時間: 有交互作用項的 Cox 模型	187
6.4	Gehan 急性白血病臨床試驗: 比較 Cox 模型在處理相同事件時間之不同計算方法	195
6.5	美國退伍軍人醫院肺癌臨床試驗: 以逐步選取法建立模型之結果 .	201
6.6	末期卵巢癌雙盲隨機臨床試驗: 變數說明	207
6.7	惡性黑色素瘤存活分析: 變數說明	209
6.8	惡性腦神經膠質瘤臨床試驗: 變數說明	211
6.9	淋巴結陰性乳腺癌隨機臨床試驗: 變數說明	213
7.1	Gehan 急性白血病臨床試驗研究: 型 (1) 存活函數估計	227
7.2	Gehan 的臨床試驗: 型 (2) 與型 (3) 存活函數估計	228
7.3	骨惡性肉瘤的存活分析: 變數說明	232
8.1	白血病患者的白血球數與死亡時間: 標準的 Cox PH 模型與分層 Cox PH 模型分析的參數估計	238
8.2	白血病患者的白血球數與死亡時間: 分層分析與非分層分析所估計的存活函數	239
9.1	AIDS 與時間相依共變數的結果	244
9.2	血壓與時間相依共變數的分析結果	245

9.3	計數過程與時間相依共變數資料輸入	254
9.4	計數過程與時間相依共變數的資料編碼	254
9.5	計數過程與多個時間點的重覆測量值之資料編碼	255
9.6	史丹佛心臟移植研究資料: 變數定義	256
9.7	史丹佛心臟移植資料: 10 位病患詳細資料	257
9.8	史丹佛心臟移植資料; 時間相依 Cox 模型結果	263
9.9	UIS 藥物濫用和預防 AIDS 臨床試驗: 變數說明	267
9.10	末期腎臟病腎臟移植存活研究: 變數說明	269
9.11	末期腎臟病腎臟移植存活研究: 10 位個體部分資料	269
9.12	腎臟移植後喪失腎臟移植體功能的事件時間	270
10.1	原發性膽汁肝硬化	289
11.1	風險分數分割結果與存活預測	307
11.2	醫學檢驗與假說檢定	308
11.3	原發性膽汁肝硬化存活分析: Cox 存活預測模型	312
11.4	晚期卵巢癌長期追蹤: 變數說明	315
11.5	基因表現與乳癌存活分析 (NKI): 變數說明	317
11.6	子宮頸癌與腫瘤指標: 變數說明	319
12.1	安養中心老人死亡時間: 部分資料	322
12.2	安養中心老人死亡時間: 變數說明	322
12.3	AIDS 潛伏期時間: 變數說明	330
12.4	AIDS 潛伏期時間: 部分個體資料	330
12.5	所有個體以右截略資料為基礎的 AIDS 潛伏期時間分配估計值	332
13.1	乳癌病患乳房外貌惡化變縮的時間 (單位: 月)	343

13.2	以所有乳癌病患乳房外貌美觀惡化變縮之區間設限的資料為根據, 估計其存活機率	344
13.3	愛滋病毒抗體檢測調查: 變數說明	347
14.1	糖尿病視網膜病變臨床試驗 (DRS): 部份資料	351
14.2	糖尿病視網膜病變臨床試驗 (DRS): 變數說明	351
14.3	膀胱癌臨床試驗: 部份資料	352
14.4	膀胱癌臨床試驗 (Bladder): 變數說明	352
14.5	群聚共變數同義詞	353
14.6	Gehan 急性白血病臨床試驗: 配適一般 Cox 模型, 比較獨立的參數變異數估計, 與考慮配對樣本之三明治估計式	356
14.7	多變量存活分析可能之組合	357
14.8	有序, 無序資料與分層邊際模型組合	360
14.9	大腸癌臨床試驗存活分析: 變數說明 I	363
14.10	大腸癌臨床試驗存活分析: 變數說明 II	364
15.1	DRS 資料之群聚邊際模型: 一般 Cox 模型與加入變異數三明治估計式	368
15.2	血液透析導管感染時間分析: 變數說明	371
16.1	DRS 資料: 比較分層 Cox 模型與其他邊際模型	377
16.2	DRS 資料: 比較脆弱模型	397
16.3	配對小老鼠存活實驗: 變數說明	401
17.1	AG, PWP 邊際模型與計數過程資料輸入	411
17.2	有序, 無序資料與分層邊際模型組合	416
17.3	三種常用的有序多重事件存活分析之資料建構	417
17.4	Bladder 原始資料: AG 與 PWP 資料輸入配適一般模型	419

17.5 Bladder 資料: 比較 AG, WLW, PWP 廣義邊際模型	420
17.6 Bladder 資料: 比較 WLW, PWP 模型之治療與分層交互作用分析 . .	421
17.7 Bladder 資料: 比較考慮第一次復發與多重復發事件	422
17.8 兒童氣喘多重復發臨床試驗: 變數說明	424
17.9 囊狀纖維化臨床試驗: 肺功能惡化次數表	426
17.10 囊狀纖維化臨床試驗: 變數說明	427
18.1 膀胱癌臨床試驗 (Bladder) 資料: 比較 AG 邊際模型與脆弱模型 (I) .	434
18.2 膀胱癌臨床試驗 (Bladder) 資料: 比較 AG 邊際模型與脆弱模型 (II) .	435
18.3 膀胱癌臨床試驗 (Bladder) 資料: 比較 AG 邊際模型與脆弱模型 (III)	435
18.4 膀胱癌臨床試驗 (Bladder) 資料: 比較 AG 邊際模型與脆弱模型 (IV)	436
18.5 膀胱癌臨床試驗 (Bladder) 資料: 比較 AG 脆弱模型	438
18.6 膀胱癌臨床試驗 (Bladder) 資料: 比較 WLW 脆弱模型	439
18.7 膀胱癌臨床試驗 (Bladder) 資料: 比較 PWP 脆弱模型	440
18.8 慢性肉芽腫病嚴重感染時間分析 II: 變數說明	443
19.1 社區感染肺炎住院存活: 變數說明	454
19.2 社區感染肺炎住院存活: 部分資料	454
19.3 社區感染肺炎: 積累發生函數 $\widehat{I}_k(t)$	455
19.4 CAP 資料: 競爭風險迴歸模型	467
19.5 何杰金氏淋巴瘤存活分析: 變數說明	469
20.1 Cox-Markovian Model 多重狀態: 資料輸入	483
20.2 中老年人髓關節骨折: 變數說明	484
20.3 中老年人髓關節骨折: 部份資料	484
20.4 中老年人髓關節骨折: 多重狀態模型	487
20.5 中老年人髓關節骨折: 時間相依共變數模型	488

20.6 兒童氣喘嚴重程度與控制	489
20.7 骨髓移植存活分析：變數說明	491
21.1 簡單可加性危險模型	502
21.2 可加性危險模型	504
21.3 可加性危險模型特例	505

圖目錄

1.1 累積發生率與盛行率: ● 起始追蹤, ✕ 死亡, ○ 存活	4
1.2 發生速率與人-年 (Incidence Density and Person-Year): ● 初始追蹤, ✕ 死亡, ○ 存活.	6
1.3 日曆時間 (Calender Time): 從 1980 到 1990 追蹤 7 位有肺癌的個體 情況. (● 起始追蹤, ✕ 死亡, ○ 存活.)	11
1.4 病患時間: 從 1980 到 1990 追蹤 7 位有肺癌的個體情況. (● 起始追蹤, ✕ 死亡, ○ 存活.)	12
2.1 不同型式的危險函數	35
2.2 常見的設限時間形式: ● 起始追蹤, ✕ 死亡, ○ 存活.	41
2.3 同時存在左截略與右截略觀察資料	48
3.1 不同危險速率的指數型存活函數	57
3.2 不同形狀的韋伯存活函數及危險函數	59
3.3 DMBA 資料的非參數之 Kaplan-Meier 存活曲線與指數存活曲線	78
4.1 時間軸與分割之時間區間	94

4.2	VGH 第一期非小細胞肺癌手術後存活研究的生命量表: (a) 存活函數 $S(t)$, (b) $-\log(\text{存活函數})$, (c) $\log(-\log(\text{存活函數}))$, (d) 機率密度函數, 以及 (e) 危險函數	101
4.3	Kaplan-Meier 存活曲線: 時間軸與分割之時間區間	106
4.4	Kaplan-Meier 存活曲線	109
4.5	急性白血病臨床試驗中 21 位孩童接受 6-MP 治療: Kaplan-Meier 存活曲線; (左) 存活曲線, (右) 包含對稱信賴區間的存活曲線.	114
4.6	復發惡性腦星狀細胞瘤的 Kaplan-Meier 存活曲線	120
4.7	VGH 第一期非小細胞肺癌手術後存活研究: Kaplan-Meier 存活曲線及各種信賴區間帶	121
5.1	Gehan 的急性白血病研究: 比較兩個樣本之存活曲線	128
5.2	Gehan 急性白血病研究的存活曲線及信賴帶	136
5.3	F98-神經膠質瘤老鼠動物實驗: 比較三組治療之的存活曲線及累積危險函數曲線	143
5.4	美國退伍軍人醫院肺癌臨床試驗之存活曲線: 左圖為比較兩種治療存活率; 右圖為比較四種細胞存活率	145
5.5	美國退伍軍人醫院肺癌臨床試驗: 以 4 級細胞型態作為分層下的 2 組治療方式之存活曲線	146
6.1	白血病患者的白血球數與死亡時間資料的存活函數: (左) 以 Kaplan-Meier 方法為基礎; 與 (右) $\log(-\log(\text{survival}))$	185
7.1	Gehan 急性白血病臨床試驗研究: 比較 Kaplan-Meier 與 Cox PH 模型下的形式 (3) 估計存活曲線	229
7.2	美國退伍軍人醫院肺癌臨床試驗: 基於 Cox PH 模型, Karnofsky 分數 (kps) 在 60 且在 4 級細胞型態之下的母體估計存活曲線.	230

8.1 白血病患者的白血球數與死亡時間: 分層分析與非分層分析所估計的存活曲線	238
9.1 移植效應: 移植視為時間相依共變數的危險函數 (Cox, 1984)	247
9.2 史丹佛心臟移植資料, 使用時間相依共變數 Cox 模型之存活曲線估計: (左): 兩個模擬世代; (右): 時間相依共變數個體.	265
10.1 Cox-Snell 殘差對殘差的估計累積危險率圖:(左) 整體; 與 (右) edema 的三個群組.	290
10.2 Cox-Snell 殘差對來自 edema 分層 Cox 模型殘差的估計累積危險率圖: (左) 整體; 與 (右) edema 的三個群組.	291
10.3 虛無模型的平賭殘差對共變數圖形.	292
10.4 加入共變數的平賭殘差對共變數圖形.	293
10.5 第 k 層 Kaplan-Meier 存活曲線的分層估計式 $\widehat{S}_k(t), \log(-\log(\widehat{S}_k(t))) = \log(\widehat{H}(t Z=k))$ 對 t 的圖形.	294
10.6 scaled Schoenfeld 殘差對時間 futime 繪製平滑曲線圖形.	295
10.7 離群值的 Deviance 殘差分析.	296
10.8 影響點的 Dfbeta 分析.	297
11.1 Receiver Operating Characteristic Curve (ROC curve)	307
11.2 訓練資料集: ROC 與 AUC	312
11.3 測試資料集: ROC 與 AUC	313
11.4 訓練資料集與測試資料集: AUC	313
12.1 (左) 危險集合數量為年齡函數; (右) 給定進入年齡 > 68 , 估計之條件存活機率	325
12.2 忽略截略時間 (從出生後或從進入安養中心後計算存活時間) 及考慮截略時間 (全部, 男性及女性) 的存活機率, 時間軸從 0 開始.	326

12.3 忽略截略時間 (從出生後計算存活時間) 及考慮截略時間 (全部, 男性及女性) 的存活機率, 時間軸從 68 歲開始.	327
12.4 紿定感染時間小於 8^+ 年, AIDS 潛伏期時間時間之條件機率, $\widehat{F}_X(x) = P(X \leq x X \leq \tau = 8^+)$, 的估計曲線, 其中, $x = \tau^+ - r$	333
13.1 Turnbull intervals: 10 位個體的區間設限資料之時間分割.	340
13.2 以所有乳癌病患乳房外貌美觀惡化之區間設限的資料為根據, 估計其存活曲線	345
16.1 比較脆弱模型之條件危險函數與邊際危險函數	382
16.2 在脆弱模型之下, 比較條件危險函數與邊際危險函數的風險比. 隨著時間增加, 邊際危險函數之危險比值趨近於 1, $\lim_{t \rightarrow \infty} HR^M(t) \rightarrow 1$. . .	383
17.1 三種常見之有序多重存活分析的邊際模型 (Therneau 與 Grambsch, 2000)	409
19.1 競爭風險模型: 第 k 事件原因別危險函數 (CSH) $\lambda_k(t)$, $k = 1, 2, \dots, K$. .	448
19.2 CAP: 競爭風險模型	455
19.3 CAP: 在競爭風險模型下的累積發生函數估計.	455
19.4 子分配危險函數 Subd-H 的風險集合計算 (Lau, Cole 與 Gange, 2009) .	459
19.5 CAP: 累積發生率函數 (CIF) PSI 4 vs 5	463
20.1 多重狀態模型: (a) 傳統存活模型, 2 狀態; (b) 單一方向或復發事件 (Fertility Model 或 Repeated Events), 4 狀態; (c) 競爭風險, 3 狀態; (d) 多重狀態.	472
20.2 單一方向疾病死亡模型	477
20.3 Nelson-Aalen 估計原因別危險函數: 狀態 0 = 第 1 次骨折手術, 狀態 1 = 第 2 次骨折, 狀態 2 = 死亡.	485

20.4 Aalen-Johansen 估計轉移機率: 狀態 $0 =$ 第 1 次骨折手術, 狀態 $1 =$ 第 2 次骨折, 狀態 $2 =$ 死亡.	486
20.5 Landmark 分析轉移機率: 狀態 $0 =$ 第 1 次骨折手術, 狀態 $1 =$ 第 2 次 骨折, 狀態 $2 =$ 死亡.	486
21.1 簡單可加性危險模型的累積迴歸函數	503
21.2 比較 Nelson-Aalen 估計式與 Aalen 可加性危險模型: 累積危險函數 估計	503
22.1 布瓦松過程與布朗運動	511
22.2 計數過程示意圖	513

